Available online at www.prace-ri.eu

Partnership for Advanced Computing in Europe

Security in an evolving European HPC Ecosystem

Dirk Pleiter[a*], Sebastien Varrette[b], Ezhilmathi Krishnasamy[b], Enver Özdemir[c], Michał Pilc[d]

[a]*PDC Center for High Performance Computing, KTH, Sweden*
[b]*Department of Computer Science, Université du Luxembour, Luxembourg*
[c]*Informatics Institute, Istanbul Technical University, Turkey*
[d]*Poznan Supercomputing and Networking Center, Poznan, Poland*

Abstract

The goal of this technical report is to analyse challenges and requirements related to security in the context of an evolving European HPC ecosystem, to provide selected strategies on how to address them, and to come up with a set of forward-looking recommendations. A key assumption made in this technical report is that we are in a transition period from a setup, where HPC resources are operated in a rather independent manner, to centres providing a variety of e-infrastructure services, which are not exclusively based on HPC resources and are increasingly part of federated infrastructures.

1 Introduction

In recent years, security incidents did have a severe impact on the availability of several high-end HPC systems throughout Europe. Securing HPC systems can be expected to become more challenging due to increasing complexity of the ecosystem, in which these systems are being operated. HPC systems are increasingly often operated in conjunction with other compute resources, e.g. private cloud instances, as well as different storage resources. Additionally, the e-infrastructure services, through which these resources are accessed, are becoming federated. This evolution of the HPC ecosystem is one of the reasons for rethinking the approaches to managing security and putting information security management systems in place. Also, the changing risk assessment due to an increased risk of cyber-attacks as well as the need for a higher level of protection, e.g. in the context of the processing of sensitive data, has to be taken into account.

We follow in this technical report the language used in the EU regulation 2019/881 [EU2019] and use the term security (or cybersecurity) to refer to any activities necessary to protect network and information systems, the users of such systems, and other persons affected by cyber threats.

This technical report is organised as follows: In Section 2 we describe the expected evolution of the European HPC ecosystem and provide a reminder of the main types of threats. We continue with a discussion of a selected set of approaches to improve security in Section 3 and an analysis of how security standards can be leveraged in Section 4. This is followed by a summary and a set of recommendations in Section 5.

This technical report is part of a series of reports published in the Work Package "HPC Planning and Commissioning" (WP5) of the PRACE-6IP project. The series aims to describe the state-of-the-art and mid-term trends of the technology and market landscape in the context of HPC and AI, edge-, cloud- and interactive computing, Big Data and other related technologies. It provides information and guidance useful for decision makers at different levels: PRACE aisbl, PRACE members, EuroHPC sites and the EuroHPC advisory groups "Infrastructure Advisory Group" (INFRAG) and "Research & Innovation Advisory Group" (RIAG) and other European HPC sites. Further reports published so far on the PRACE webpage[1] cover "State-of-the-Art and Trends

* Corresponding author email address: pleiter@kth.se
[1] https://prace-ri.eu/infrastructure-support/market-and-technology-watch/

for Computing and Network Solutions for HPC and AI", "Data Management Services and Storage", "Edge Computing: An Overview of Framework and Applications", "Quantum Computing - A European Perspective" and "User Requirements influencing HPC Technologies".

2 Background

2.1 Evolving HPC infrastructures

In this subsection, we consider some trends in HPC that have a strong impact on future approaches to security.

In the past, the primary role of HPC centres was to operate one or more supercomputers and to provide access to this resource. In the future, we expect these centres to transform to service providers, where the latter are based on different types of underlying compute and storage resources. While this will continue to include supercomputers, an increasing number of centres have also deployed, e.g. on-premise private cloud instances. Most of the new EuroHPC pre-exascale and petascale systems do include such instances. These services started to become harmonised and federated, e.g. in the context of the Fenix initiative[2]. Meanwhile the trends sketched here have meanwhile become part of the future EuroHPC strategy, as EuroHPC added "HPC Federation and Services" as one of the five pillars of future activities.[3]

With these architectural changes, HPC centres respond to changing user needs as well as those of emerging new science and engineering domains, which do need HPC resources for their research. A first trend to highlight are efforts towards establishing domain-specific platform services that allow end-users to run HPC workflows through open portal services. Such platform services are being implemented among others by projects like the Human Brain Project[4] or different ESFRIs like those organised in the European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE)[5] or the ESFRI for life-science information Elixir[6]. Furthermore, we observe a growing demand for using HPC infrastructures for storing and processing sensitive data like personal data, where compliance with the EU General Data Protection Regulation (GDPR) is mandatory [EU2016]. Finally, HPC infrastructures have to become ready for more support of users coming from private organisations, e.g. industry or SMEs.

These trends require rethinking our approach to security in the context of HPC infrastructures for multiple reasons. Firstly, workflows will run on HPC resources, where access is restricted, but are connected to and triggered through relatively openly accessible platform services like web portals. Secondly, these workloads and platform services need to be able to run on top of federated e-infrastructure services provided by different sites. Therefore, a harmonisation of the security levels provided by different sites is required. Finally, processing and storing of sensitive data as well as service offerings for users from private organisations lead to higher demand for security to ensure the protection of data and to ensure a high level of confidentiality for instance to protect trade secrets.

The HPC community can benefit from the important role that security is playing for commercial suppliers of Cloud e-infrastructures, where the commercial impact of security breaches is more obvious. Security incidents may not only result in a loss of customers, but liability obligations may have severe financial consequences.

2.1 Main concepts and threats

HPC facilities are exposed to threats inherent to digital communications when interacting with such large-scale computing systems. The formalization and mitigation of these threats are widely addressed in the literature tied to the cryptology and network security domains [Dumas2015]. The main functionalities aimed to be guaranteed in this context are traditionally summarized using the acronym CAIN: Confidentiality, Authentication, Integrity, Non-repudiation.

In digital communication systems, data ***confidentiality*** means that data is only accessible to a specific set of users. Confidentiality has become more important as more sensitive data is being processed, e.g. in the context of life sciences. Typical measures to support confidentiality are to provide mechanisms for restricting access to data, e.g. by means of access control lists (ACL), or by encrypting data using keys. The encryption can be symmetric or

[2] https://www.fenix-ri.eu

[3] European Commission, "Equipping Europe for world-class High Performance Computing in the next decade", SWD(2020) 179 final, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020SC0179&rid=9.

[4] See, e.g. EBRAINS Simulation services offering (https://ebrains.eu/services#category2)

[5] https://projectescape.eu/

[6] https://elixir-europe.org/

asymmetric. Symmetric key algorithms employ a single secret key which is used for both encryption and decryption, while asymmetric key algorithms, like public-key algorithms, use different keys for encryption and decryption. In the context of HPC, encryption can be challenging for different reasons. The performance impact of data encryption and decryption can be significant, particularly in extreme-scale data sets. Furthermore, there are no established and widely deployed solutions for distributing keys.

Authentication is the process where an entity (the Principal) is proving its identity to another entity (the System) that must be able to detect identity theft. This is a challenging property for federated authentication services foreseen to become standard within the European digital ecosystem. Any service provider has to be able to manage the identity of individuals (staff members, platform users, business clients) or systems and cope with the threat of identities being compromised.

Integrity of data is about ensuring data to be never altered or partially removed. With the increased amount of data stored in HPC centres even the risks related to unintended data corruption due to failures of the storage technologies is increasing. For being protected against data integrity threats it might be necessary to keep hash functions in a separate safe location. Also providing the option of keeping redundant copies of data may help preserve integrity of data.

Non-repudiation is required to provide protection against an individual falsely denying having performed a particular action. It may, in particular, involve the capability to determine whether a given individual took a particular action such as creating information, sending a message, approving information, or receiving a message [NICCS2021].

The Cloud Security Alliance (CSA), which is an organisation that works on defining and raising awareness of best practices related to security in the area of cloud computing, performed a survey a few years ago among experts to compile opinions on the most important security issues within cloud computing [CSA2016]. The identified threats do also apply to HPC data centres. As the most important security concern, the CSA report lists data breaches, i.e. incidents that lead to sensitive, protected or confidential information to be released or stolen. Next, insufficient identity, credential, and access management are listed, followed by insecure interfaces and APIs.

3 Approaches to improving security

In this section, we collect a selected set of approaches to improve security in the context of infrastructures that comprise HPC resources.

3.1 Access and identity management

3.1.1 Use of cryptographic protocols and SSH

The users of HPC centres access and use supercomputer facilities remotely. Securing such communications over an unsecured network relies on cryptographic protocols. The fundamental objective of cryptography is to enable two entities, traditionally called Alice and Bob, to communicate through an insecure channel in such a way that any opponent, Oscar, having access to the information circulating on the communication channel is not able to understand what is exchanged [Dumas2015].

There exist two main types of cypher schemes ensuring the confidentiality of the exchanged messages: secret key (or symmetric) cryptography which mimics perfect secrecy where the same key is a shared secret between the sender and the recipient and is used for both encryption and decryption of messages. In such a scheme, there is only one key that each user utilizes; therefore, this key has to be already known to all users. The messages to be sent to the other party are first encrypted by this key. When the encrypted text or cypher text is transmitted to other users, they decrypt the cypher text with the same key and recover the original messages. Symmetric key cryptosystems can be almost secure and efficient in terms of computation time. However, before using such a scheme, how can the key be distributed to all communicating parties?

This is the place where public key (or asymmetric) cryptography is needed: it eases key exchanges (typically through key exchange protocols) since no secret information is shared a priori among the involved protagonists. In the public key encryption method, unlike symmetric key algorithms, each communicating party has 2 different keys; one public key and one private key. Encryption is done with the user’s public key and decryption is only performed with the same user’s private key. The sender encrypts the message to be transmitted with the public key of the receiver. The public key is known to everyone but it does not violate the security of communication as the sent text can only be decrypted with the recipient's private key. A symmetric key algorithm is much less costly than a public key algorithm.

These concepts are implemented to secure the interactions with HPC centres, from the Secure Shell protocol (SSH) guiding the remote access to the supercomputers and described in the next section, to SSL protecting the web portal and HPC management services. For instance, SSH employs a public key algorithm to authenticate users and to distribute a shared key for symmetric key encryption. Thus the user's side SSH application encrypts the username and (the hash value of) the password with the public key of remote servers and in this case, the only way to decrypt the cypher requires having the private key of the remote server. Once the remote user is authenticated by the host, the distribution of the shared symmetric key is also performed via a public key algorithm. In other words, the shared key is encrypted with the recipient's public key then the cypher text is sent to the intended receiver.

3.1.2 Improving security of SSH-based access

By default, access to HPC facilities is enabled through Secure Shell (SSH) communication and encryption protocol (version 2). The PRACE service catalogue does foresee SSH as a "core service". This encrypted network protocol is used to log into another computer over an unsecured network, to execute commands in a remote machine, and to move files from one machine to another in a secure way. It is also used as the default medium to secure communications to remote servers with OpenSSH[7] , the most popular implementation used on nearly all systems. Most servers related to an HPC system have an SSH daemon running, for instance to allow system administrators to connect and manage the system.

Since SSH has such an important function on the HPC ecosystem, and as firewalls are often opened up to allow traffic, proper hardening of the service is needed. Nowadays, this protection is handled according to three main axes:

- Proper security hardening of the SSH configuration to reduce known weaknesses, which can, in particular be achieved through updates of /etc/ssh/sshd_config. Among the traditional recommended configuration settings, the following changes should be enforced:
    - DNS hostname checking
    - Disabling the password-based authentication and allow public key authentication (as using public key authentication is considered much safer and less prone to brute-force attacks). The way SSH handles the keys and the configuration files is illustrated in Figure 1:

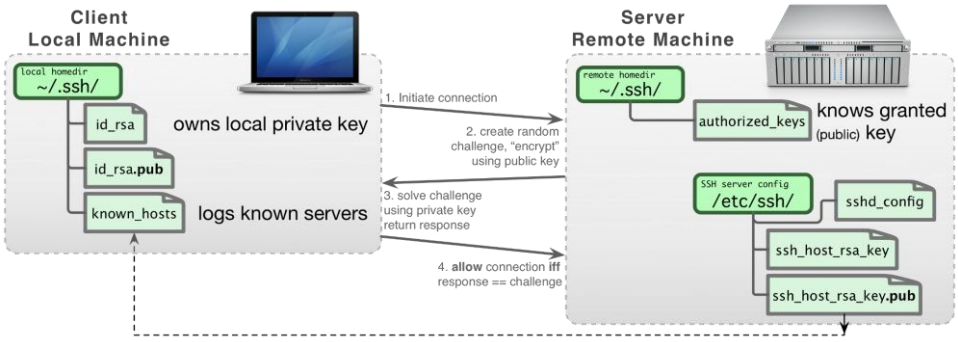

Figure 1: Illustration of the way SSH handles the keys and the configuration files

    - Disabling the root login (direct root logins may result in bad accountability of the actions performed by this user account). If possible, i.e. when not all users should have access to the system, it is advised to explicitly white-list the allowed users and groups of users, using the "default deny" principle
    - Restricting the number of authentication attempts
    - Restricting the cryptographic cyphers, key exchange algorithms and Message Authentication Code (MAC) functions to only the most secure ones recommended by NIST
    - If possible, the default listening port for the SSH service should be changed

[7] https://www.openssh.com/

References and resources that include examples to define highly secured configurations are provided in the appendix.

Additional traditional restrictions should be implemented. This can be facilitated by traffic filtering. An easy mechanism provided by the SSH protocol is to make the use of a from="pattern-list" clause in a list of authorised keys mandatory.

Furthermore, the usage of SSH configuration scanners and security tools are encouraged to be performed in a regular (and automated) way. This includes:

- Lynis[8], an open source security tool designed as a security scanner and compliance auditing tool. It can detect vulnerabilities and configuration flaws, but also provides recommendations for an in-depth audit and continuous improvement
- ssh_scan[9], a SSH configuration and policy scanner designed by the Mozilla foundation and thus kept up-to-date with the latest security hardening best practices.
- ssh-audit[10], while slightly outdated, allows to perform a test on a selected set of remote targets to analyse the responses it receives.

- Enabling Firewall to restrict and monitor SSH traffic, coupled with protective tools against brute-force attacks such as Fail2ban[11].
- Consider general system hardening by enabling Security-Enhanced Linux (SELinux) [SELinux14] is also recommended. SELinux is a security architecture for Linux systems designed to finely control and define access controls for the applications, processes, and files on a system. It uses security policies, which are a set of rules that tell SELinux what can or can't be accessed, to enforce the access allowed by a policy.

Finally, instead of allowing file-based authorised keys under the control of users, using a centralized Identity Management (IdM) service (yet potentially distributed and replicated across multiple servers to allow for high availability) could also be considered. Such a service is used to create identity stores, centralized authentication, domain control for Kerberos and DNS services, and authorization policies. It is a central component within HPC sites to allow for federated authentication services as expected in reference European projects and initiatives as PRACE, EuroHPC, Fenix[12], EOSC or even Eduroam. IdM services are traditionally handled by middleware systems such as FreeIPA and Redhat IdM, 389 Directory Server, Microsoft Active Directory, OpenLDAP. Within a HPC facility, Redhat IdM[13] or IPA[14] are native and reference Linux-based frameworks addressing this need. It offers a more secure and robust environment for handling user identity, roles and credentials (including SSH key pairs) than what was possible in the past with directory services and LDAP.

In practice, the System Security Services Daemon (SSSD) interacts as a client component of the centralized IdM service deployed within the HPC facility to handle in a transparent, consistent and secure way the authentication service on a given host based on (Host-Based Access Control) HBAC rules. It also caches the information stored in the remote directory server to provide identity, authentication and authorization services to the host machine (login server, HPC head or compute server, web or cloud portal etc.). This comes with several advantages:

- Novel secure authentication methods and schemes (Multi-Factor authentication (MFA), OICD etc.) can be transparently enabled to serve the full user community
- The potential overhead induced by federated authentication services when dealing with the trust delegation protocols and verifications can be offloaded to dedicated servers

3.2 Improving the network security architecture

HPC encompasses advanced computation on parallel systems, enabling faster execution of highly compute-intensive tasks which heavily rely on interconnect performance. For this reason, the main high-bandwidth low-latency network of an HPC facility relies on the dominant interconnect technology in the HPC market, i.e. InfiniBand (IB) or HPE/Cray Slingshot. However, as these networks are internal they are less affected by security

[8] https://cisofy.com/downloads/lynis/

[9] https://github.com/mozilla/ssh_scan

[10] https://github.com/arthepsy/ssh-audit

[11] https://github.com/fail2ban/fail2ban

[12] https://www.fenix-ri.eu

[13] Redhat Idm: https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/7/html/linux_domain_identity_authentication_and_policy_guide/introduction

[14] FreeIPA: https://www.freeipa.org

threats. Additionally, technology solution providers offer advanced monitoring infrastructures (e.g. Nvidia/Mellanox Unified Fabric Manager (UFM)[15]) that allow to assess the security and the performance of these networks.

The main security concerns are tied to the IP networks available at all HPC facilities. To limit the risk of network intrusion and to control the access to sensitive information, a network segmentation and segregation policy must be defined and implemented. The key features expected by the network organization are:

1. Scalability, i.e. support of
    - Thousands of computing and data storage elements which may be physically distributed across different racks and/or server rooms
    - Many virtualised systems per computing element
    - Hundreds of services, both external (Internet-facing) and internal
    - Direct connection to hosted HPC high-throughput equipment
    - Random user data access and throughput patterns during the day (majority) and night (minority)
    - Constant administrative data throughput patterns during the night (on site / off site backups)
    - Service level agreements (SLA) with internal/external users
2. Isolation from the hosting entity's internal network. This is realised by a separation of the different flows and streams within dedicated Virtual Local Area Network (VLAN), with strict policies enforced at all levels:
    - Access network, aiming for DMZ (demilitarized zone) services i.e. accessible by users on one side, linking to the data/Infiniband networks on another side
    - Production network: meant for user-level data transfer and Internet access, in-band management
    - Management network: meant for out of band management of base hardware by the HPC operation team
    - Infiniband network: user-level very-high bandwidth, very-low latency data transfer

This assumes systematic rules for IP addressing being in a place such that there is no overlap with a range reserved by the hosting entity. In the physical implementation of the network, a systematic network cabling policy should also be enforced, which foresees identical and unique labels on both ends of each cable.

Figure 2 illustrates the implementation of an IP network for a large-scale HPC and Big Data analytics infrastructure following secure best-practices in terms of VLAN structure.

It is recommended to organize such a network as a 2-layer topology: one upper level (Gateway Layer) with routing, switching features, network isolation and filtering (ACL) rules and meant to aggregate only switches and routers. As can be seen, this is where the interfaces to the local organization network, the outside world (i.e. public internet), as well as external public or private partners (including HPC centres or other members of a federating structure) takes place. The bottom level (Switching Layer) is typically composed of core switches as well as the TOR (Top-of-rack) network equipment, meant to interface the HPC servers and compute nodes. Both layers define at least three isolated VLANs that allow structuring the different flows and streams and thus with different security ACLs: the sensitive DMZ VLAN for external accesses, either to the login nodes or to the storage equipment, the production VLAN serving user applications with the computing facility, the Management VLAN for the corresponding tasks of the operational team responsible for maintaining and administering the system. In addition, non-routed VLANs are recommended: one to serve applications that do not support natively the fast interconnect technology in place (Infiniband for instance) through an IP emulation layer in addition to the production network. Then cloud modules would benefit from a dedicated VLAN offering an overlay network isolating the virtual machines and services deployed within this module. Secure implementations as KVlan within Grid5000[16] or Chameleon[17] illustrate the capabilities of such VLANs.

[15] https://www.nvidia.com/en-us/networking/infiniband/ufm/

[16] https://www.grid5000.fr/

[17] https://chameleoncloud.readthedocs.io/

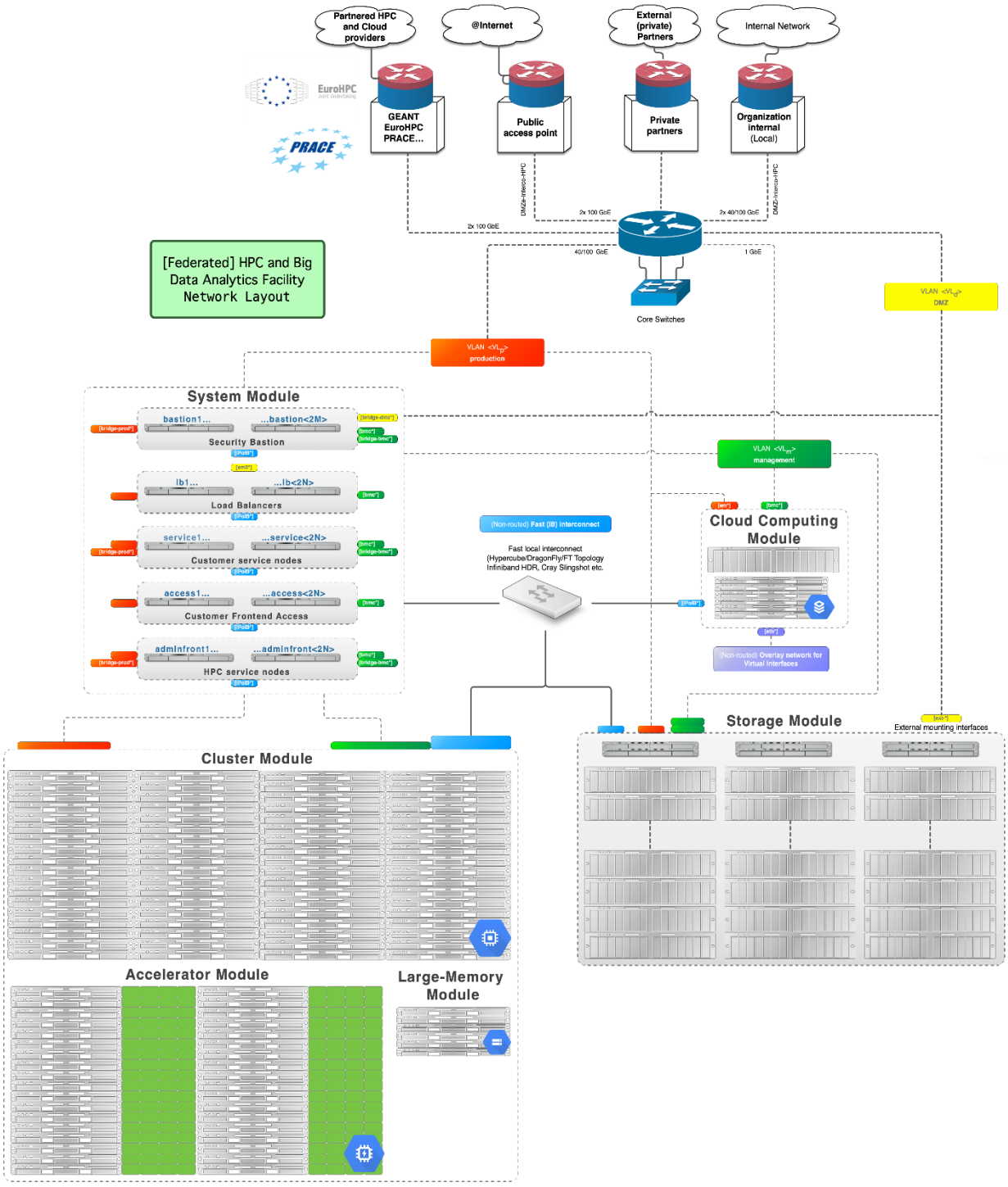

Figure 2: Schematic view of IP network for a large-scale HPC and Big Data analytics

3.3 Cybersecurity and Big Data Storage protection

With the embedded storage capacities of large-scale computing facilities, novel security challenges appear with, on the one hand the emerging paradigm of Open Science enabling an easier access to expert knowledge and material, and on the other hand the necessary compliance to the EU's General Data Protection Regulation (GDPR) [EU2016].

The interactions occurring during data processing activities are covered in the literature, for instance in [Paseri2021] and illustrated in Figure 3.

This illustrates the different types of data processing steps to be protected:

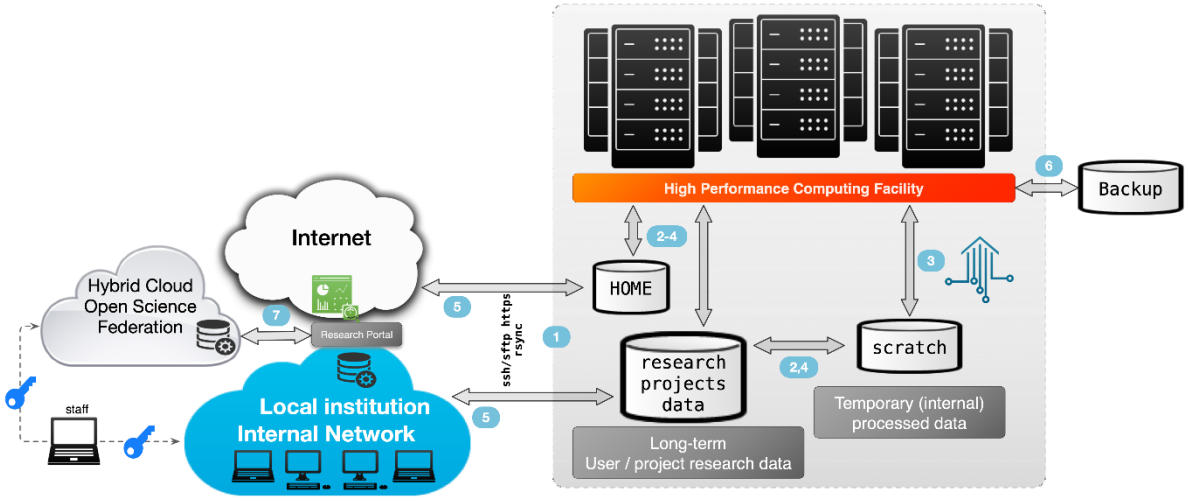

Figure 3: Illustration of data analytics workflow in a federated environment.

1. Data transfer of the input data towards the long-term storage area
2. Pre-processing phase to prepare the data for research analysis (this may include partial or total data transfer towards the internal scratch area)
3. Job processing on the computing facility, generating both intermediate and final data components
4. Post processing phase to derive scientific results from the processed data: this typically includes the creation of a metadata catalogue allowing to index and quickly recover scientific data
5. Data transfer of the output data towards external resources (for instance a laptop of an external server).
6. Archiving of the results, and backup of the long-term storage area
7. Data replication and synchronization may happen in a federated environment; in parallel, data sharing can be performed in the context of research collaborations; this includes live processing and access by external stakeholders

One of the major security challenges foreseen is represented by the complex tracking of data movements done in the steps 2 to 4 of the above figure. Parallel and distributed file systems used in HPC environments are not yet fully able to account and log internal data movements. More precisely, changelogs-based auditing capabilities relevant for the GDPR compliance are only featured starting with recently released versions of Lustre (2.11) and IBM Spectrum Scale (5.0). With regards to the other type of data transfer performed in the considered workflows (i.e. steps 1, 5 and 7), other accountability and monitoring mechanisms can be implemented to fulfil this constraint.

3.4 Introducing monitoring and intrusion detection systems

An intrusion detection system (IDS) is a device or software application that monitors a network or system for malicious activity or policy violations. Intrusion detection systems fall into one of three categories: Host Based Intrusion Detection Systems (HIDS) which typically monitors the integrity of important operating system files, Network Based Intrusion Detection Systems (NIDS) which analyses incoming network traffic to detect suspicious activities, and hybrids of the two. It is also possible to classify an IDS by detection approach. The most well-known variants are signature-based detection (recognizing bad patterns, such as malware) and anomaly-based detection (detecting deviations from a model of "good" traffic, which often relies on machine learning).

IDS in HPC environments is rather challenging since such large-scale facilities tend to have very distinctive modes of operation or be used for very distinctive purposes depending on the user profile. However, they are necessary to detect cyber-attacks against these systems, i.e. unauthorized access with a malicious intent to steal sensitive documents, compromise networks, vandalize the resources or use the resources for further malicious actions. Supercomputers across Europe were for instance recently infected (in May 2020) with cryptocurrency mining malware, forcing operators to shut the systems down to investigate the attack. Details were made public in an advisory from the European Grid Infrastructure (EGI) CSIRT about these cases, where it is claimed that compromised servers in Poland, Canada and China were used in these attacks[18]. The CERN Computer Security

[18] https://csirt.egi.eu/attacks-on-multiple-hpc-sites/

Team and other organizations identified the usage of the Kobalos malware which predates these incidents[19]. It is thus crucial to at least detect such malware within the implemented IDS systems.

To detect whether a system is compromised, also the use of machine learning methods for automated analysis of system behaviour has been explored [Peisert2017]. The idea was to search for indications where users were running unusual workloads.

To assess the current use of IDS solutions in European HPC centres, the three questions shown below were distributed among those centres that are organised in PRACE. Five HPC centres responded. For obvious reasons their identity is not disclosed here.

Question 1: What kind of IDS tools do you use?

Two out of five centres reported that they use a dedicated IDS. Both use AIDE[20], but in one case it is supplemented with additional dedicated scripts to detect unusual spikes in resource usage.

In the remaining three HPC centres alternative solutions are applied, namely:

1. IP address whitelisting on nodes connected to the public Internet implemented with a hardware firewall; in addition, all incoming and outgoing connections are logged and sent via rsyslog to a centralized logging server
2. No dedicated IDS; a hardware-based solution from Fortinet that is mainly used for establishing VPN connections but has built-in IDS functionalities.

Question 2: What is a typical configuration of your IDS? What parameters are monitored?

All responding sites that use an IDS system reported that malicious software signatures are monitored and tracked instantaneously. One of the HPC centres uses Ganglia for monitoring and Nagios to receive alerts and when CPU/memory/disk consumption exceeds predefined limits. Two sites reported that their systems are configured such that changes in the file system of management master nodes are observed. One site reported that they enabled tracking of login details (unsuccessful attempts, source IP addresses of successful and unsuccessful login).

Question 3: Does the IDS tool used at your site impact new services like OpenStack?

One response did not address this question in a clear manner. In all other responses no negative impact of IDS on services like OpenStack is reported. Two of them do not provide any explanation for this observation. In one case, the site believes that the lacking negative impact of using an IDS is because neither virtualisation nor any sophisticated network setups are being used. One site did highlight that their IDS solution is limited by the rate at which the data can be processed, which is currently 11 Gbps. They expect to have 52 Gbps in the future.

4. Leveraging security standards

In the previous sections it was shown that security issues have become more important and need to be addressed at a level that exceeds a single data centre as sites are starting to offer federated e-infrastructure services. One strategy to enhance security and establish common security levels is to leverage standards. In this section a number of such standards is reviewed.

4.1 NIST guides for conducting risk assessment

Dealing with security in the context of operating e-infrastructure services requires balancing the possible impact of threats on the one hand and the needs for openness and usability of these services as well as the costs associated with security measures on the other hand. This can be framed as a risk management process and the need for creating an environment, where decisions can be made based on an assessment of the risks. NIST [NIST2012] developed a guide for systematising the process of risk assessment. The latter should lead to an identification of all relevant threats, vulnerabilities to organizations, the impact that may occur and the probability that risks will materialise. More technical guidance on enhancing security of network and information systems is provided by NIST in a technical guide [NIST2008]. Here a number of approaches to security testing and examination are described. This includes, in particular, vulnerability validation techniques like penetration testing.

[19] https://www.welivesecurity.com/2021/02/02/kobalos-complex-linux-threat-high-performance-computing-infrastructure/

[20] https://aide.github.io/

4.2 ISO/IEC 27000 standards on information security management

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) published a set of standards related to information security management, which are known as the ISO/IEC 27000 family of standards. Starting with the ISO/IEC 27001:2018 standard a general framework for defining information security management systems (ISMS) including a vocabulary [ISOIEC2018].

In this context the most relevant standard is the ISO/IEC 27001:2013 standard [ISOIEC2013]. The standard specifies the requirements for implementing and maintaining an effective information security management system (ISMS). As such it focuses on organisational aspects and does not prescribe specific technical measures that need to be taken. For example, the standard describes how leadership is to be involved, how planning should happen, what resources to allocate for support and how to regularly evaluate the implemented ISMS.

An important aspect of this standard is the option for certification. This allows organisations to document compliance with this standard based on a successful audit by an accredited certification body.

4.3 EU Cybersecurity Directives and Regulations

In the future, stipulations provided by the European Commission related to protection of network and information systems operated in EU member states will become more important. In 2016 the EU Parliament and Council confirmed the goal of achieving a high common level of security of network and information systems within the EU [EU2016b]. It considers the member states as the actors and focuses on services like online marketplaces, search engines and cloud computing services. In 2019 the EU Parliament and Council adopted a regulation on the establishment of a European Union Agency for Cybersecurity (ENISA)[21] [EU2019], i.e. the EU is starting to become an actor. ENISA is mandated with the creation and maintenance of a European cybersecurity certification framework and to work on the necessary technical ground for more specific standards and schemes for certification based on the framework provided by the aforementioned regulation. At this point, ENISA has not published standards yet that are of particular relevance in the context of this technical report.

4.4 German Federal Office for Information Security's catalogue C5

The Federal Office for Information Security in Germany (BSI)[22] introduced in 2016 a Cloud Computing Compliance Criteria Catalogue (C5). In 2020 an updated version of C5 was published [BSI2020]. BSI is an agency of the German federal government, which is in charge of managing computer and communication security for the German government. The C5 catalogue is intended to be used by different stakeholders to use the proposed criteria for a risk assessment. It is intended to be used by cloud service providers, customers and auditors.

The catalogue defines a set of base criteria that can be used as a checklist by relevant stakeholders including the operators of network and information systems. While the C5 catalogue mainly targets (commercial) cloud providers, many of these criteria can also be used for current and upcoming HPC infrastructures. For this reason, Fenix created the "Fenix Security Measures Catalogue" [Fenix2020] based on the C5 catalogue.

The criteria are grouped with an object defined for each of these groups of criteria. The following table shows a selection of such groups of criteria that are more important in this context:

Group of C5 criteria	Objective
Organisation of information security (OIS)	Plan, implement, maintain and continuously improve the information security framework within the organisation
Security policies and instructions (SP)	Provide policies and instructions regarding security requirements and to support business requirements
Personnel (HR)	Ensure that employees understand their responsibilities, are aware of their responsibilities with regard to information security, and that the organisation's assets are protected in the event of changes in responsibilities or termination.
Asset management (AM)	Identify the organisation's own assets and ensure an appropriate level of protection throughout their lifecycle.

[21] https://www.enisa.europa.eu/

[22] https://www.bsi.bund.de

Physical security (PS)	Prevent unauthorised physical access and protect against theft, damage, loss and outage of operations.
Identity and Access Management (IDM)	Secure the authorisation and authentication of users of the Cloud Service Provider (typically privileged users) to prevent unauthorised access.
Communication Security (COS)	Ensure the protection of information in networks and the corresponding information processing systems.
Security Incident Management (SIM)	Ensure a consistent and comprehensive approach to the capture, assessment, communication and escalation of security incidents.

4.5 French National Information Systems Security Agency's catalogue

The Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI)[23] is the French counterpart of BSI. In cooperation with BSI they published in 2018 "Prestataires de services d'informatique en nuage (SecNumCloud)" [ANSSI2018]. As it is rather similar to the C5 catalogue, no specific analysis of this security catalogue has been performed.

5. Summary and Recommendations

Starting from the observation that the European HPC is evolving towards a provisioning of a federated service portfolio for serving changing and emerging user needs, current security threats and different approaches for improving security have been reviewed. Some of these approaches like the use of intrusion detection systems would not affect the openness and usability of the provided e-infrastructure services. We argued that security cannot be addressed by assuming data centres to operate in isolation. Security standards could be leveraged to harmonise security levels and security-related restrictions to the benefit of the users of the future European HPC infrastructure.

Based on our analysis, we make the following recommendations:

1. Data centres should review their security strategies within the evolving European HPC ecosystem, where an increased number of services are provided beyond providing access to a supercomputer.
2. With SSH being currently the most widely used network protocol for connecting to an HPC system, SSH configurations should be hardened for security, e.g. by enforcing DNS hostname checking, disabling password-based authentication, restricting access through white-listing methods, or by using SSH configuration scanners.
3. Use of Intrusion Detection System should be further explored as relatively few sites seem to use these today.
4. Security stands for HPC centres should be adopted at European level and should be leveraged (or even be established) to realise a common security level within a European infrastructure where services start to be federated, which would improve the usability of this infrastructure by users with specific security requirements, e.g. in the context of processing of sensitive data. The C5 catalogue from the German Federal Office for Information Security is a promising starting point, because it prescribes concrete measures.
5. The collaboration between HPC centres in Europe should be strengthened to improve the response to security incidents, which in future are even more likely to affect more than one site. Such collaboration would also allow to harmonise security measures to avoid users having to deal with different security restrictions.

[23] https://www.ssi.gouv.fr/en/

Glossary

ACL	Access Control List
CAIN	Confidentiality, Authentication, Integrity, Non-repudiation
DNS	Domain Name Service
DMZ	Demilitarized Zone
ESFRI	European Strategy Forum on Research Infrastructures
GDPR	General Data Protection Regulation
HPC	High-Performance Computing
IDS	Intrusion Detection System
ISMS	Information Security Management System
VPN	Virtual Private Network

References

[ANSSI2018] L'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), "Prestataires de services d'informatique en nuage (SecNumCloud)", June 2018, https://www.ssi.gouv.fr/uploads/2014/12/secnumcloud_referentiel_v3.1_anssi.pdf

[BSI2020] BSI, "Cloud Computing Compliance Criteria Catalogue – C5:2020", 2020, https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/ComplianceControlsCatalogue/2020/C5_2020.pdf

[CSA2016] Cloud Security Alliance, "'The Treacherous Twelve' Cloud Computing Top Threats in 2016", 2016, https://cloudsecurityalliance.org/artifacts/the-treacherous-twelve-cloud-computing-top-threats-in-2016/

[Dumas2015] J.-G. Dumas, J.-L. Roch, E. Tannier, and S. Varrette, "Foundations of Coding: Compression, Encryption, Error-Correction", Wiley & Sons, 376 pages, ISBN 978-1-118-88144-6, 2015

[Fenix2020] Fenix Consortium, "Fenix Security Measures Catalogue", August 2020, https://fenix-ri.eu/

[EU2016] EU, "Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data", 2016, https://data.europa.eu/eli/reg/2016/679/oj

[EU2016b] EU, "Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union", 2016, http://data.europa.eu/eli/dir/2016/1148/oj

[EU2019] EU, "Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)", 2019, https://eur-lex.europa.eu/eli/reg/2019/881/oj

[ISOIEC2013] ISO, IEC, "ISO/IEC 27001:2013: Information technology — Security techniques — Information security management systems — Requirements", 2013, https://www.iso.org/standard/54534.html

[ISOIEC2018] ISO, IEC, "ISO/IEC 27000:2018: Information technology — Security techniques — Information security management systems — Overview and vocabulary", 2018, https://www.iso.org/standard/73906.html

[NICCS2021] NICCS, "Cybersecurity Glossary", https://niccs.cisa.gov/about-niccs/cybersecurity-glossary (accessed on 28.09.2021)

[NIST2008] NIST, "Technical Guide to Information Security Testing and Assessment", SP 800-115, September 2008, https://doi.org/10.6028/NIST.SP.800-115

[NIST2012] NIST, "Guide for Conducting Risk Assessments", SP 800-30 Rev. 1, September 2012, https://doi.org/10.6028/NIST.SP.800-30r1

[Paseri2021] L. Paseri, S. Varrette, and P. Bouvry, "Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes," in Proc. of the EU Annual Privacy Forum (APF) 2021, 2021, vol. 12703, pp. 123–142.

[Peisert2017] Peisert, Sean. "Security in high-performance computing environments" Communications of the ACM 60.9 (2017): 72-80.

[SELinux14] Bill McCarty. 2004. “SELinux: NSA's Open Source Security Enhanced Linux”. O'Reilly Media, Inc. https://dl.acm.org/doi/10.5555/1096126

Acknowledgements

This work was financially supported by the PRACE project funded in part by the EU’s Horizon 2020 Research and Innovation programme (2014-2020) under grant agreement 823767.

Appendix

OpenSSH Server and Client secure configuration

Although most default OpenSSH settings that already good from a security perspective, we encourage further efforts for securing OpenSSH servers configurations (/etc/ssh/sshd_config) and adapt the settings according to best security practices. More specifically, we suggest HPC operational teams to follow the recommendations of the Security Assurance and Security Operations team of the Mozilla Foundation[24].

Security/Server Side TLS configuration

The configuration of TLS within all web-enabled services is a challenging task for HPC operational teams. To that end, the same team from the Mozilla Foundation propose a reference guide[25] which includes an SSL Configuration Generator[26].

[24] https://infosec.mozilla.org/guidelines/openssh.html

[25] Security/Server Side TLS: https://wiki.mozilla.org/Security/Server_Side_TLS#

[26] https://ssl-config.mozilla.org/